e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

REACH: A SEARCH ENGINE FOR A PRECISE QUERIES

Maithili Varankar*1, Tushar Prajapati*2, Akshata Chile*3, Prof. Smita Pai*4

*1,2,3student, Department Of Information Technology, Terna Engineering College, Nerul, Navi Mumbai, Maharashtra, India.

*4Prof., Department of Information Technology Engineering, Terna Engineering, Nerul, Navi Mumbai, Maharashtra, India.

ABSTRACT

Search Engine is used for searching the information about a topic. It basically retrieves data from various pages using link addresses or URLs of the visited pages and arrange it accordingly which is called indexing and gives the appropriate result according to user's demand. Information which is found is mostly stored on a computer system, such as on the World Wide Web, inside a corporate or proprietary network, or in a personal computer. Today's world various search engines are available on the internet e.g., Google, Ask.com, Yahoo, All the web, Myspace, etc. The popularity of search engines can be estimated by the fact that approximately 112 * 10 6 searches are made in a single day from one search engine alone. Based on recent studies made on the structure and dynamics of the web itself, it has been analyzed that the web is still growing at a high pace, and the dynamics of the web is shifting. More and more dynamic and real -time information is made available on the web. Our aim is to design a search engine that meets the challenges of web growth and update dynamics.

Keywords: Search Engine, Web Crawler, Bot, Query Engine.

I. INTRODUCTION

The search engine allows anyone to search for anything from anywhere. The user can request any information by passing the query in the form of keywords or category. Refers to a large database of online resources such as web pages, newsgroups, programs, photos, etc. It is a useful way of finding information on the world wide web. It will bypass content on the web based on user behavior and try to provide all the "answers" based on known content. It applies progressed information innovations to fill the information hole among people and machines. It additionally empowers both clients and machines to share and consolidate data effectively and in this manner permits machines to keenly access various information sources. Because of this the computer's ability to provide similar answers is greatly boosted. Effective searching is expected to get high quality outcomes which depend on appropriate coordinating between clear cut assets and client queries. Traditional search engine used to give precise and correct answers but in a large amount. Our main aim is to gain momentum by giving the selected results which the client is looking for.

II. LITERATURE SURVEY

Literature was reviewed from various sources, like from research papers, publications books, existing bibliographic information, and recommendations by the project panel. Various data has been obtained from these statistics.

[1] SemSearch: A Search Engine for the Semantic Web[2] Semantic Search produces answers precisely for the queries of the user by taking advantage of the availability of explicit information in the semantic web. It overcomes the problem of low barrier access for the ordinary end users. Semantic Search also deals with the complex queries along with precise and self-explanatory results. It provides a quick respons to user queries. Further the Google-like UI layer allows end users to specify queries in terms of keywords. User queries are interpreted by The Text Search Layer by finding out the particular meanings of the user keywords. The Semantic Query Layer which produces search results for user queries by translating user queries into formal queries, consists of three components which include a formal query construction engine where user queries are translated into formal queries, then a query engine that uses the generated formal queries to query the specified meta-data repository and the third one is a ranking engine where the search results are given ranks according to the degree to which they satisfy the user query. SemSearch ensures that it is effective and easy to use for the ordinary end users who are not familiar with SQL –query.

[2] Distributed Image Search (DIS): This framework [8] includes camera sensors. Clients can present inquiries by indicating a picture and the framework returns subtleties of sensors that have caught comparable pictures.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

Results are positioned and the top k generally applicable coordinating sensors/pictures are gotten back to the client. The framework upholds both consistent constant questions and chronicled impromptu questions, for which reason sensor hubs store recorded pictures. DIS could be utilized to look for places with a specific visual state. A vital restriction of the above approach is that it isn't adaptable for enormous organizations as each inquiry must be shipped off all sensors. Likewise, the pursuit is surmised as picture matching depends on highlight vectors. Albeit the framework has been intended for picture look, the issue is planned to catchphrase look, so a portion of the instruments might be relevant to different kinds of sensors that can likewise be planned to catchphrases. Content-based search and portability are upheld as sensor information are not pushed to the mediator, yet rather neighborhood calculations are performed when a new picture is caught.

[3] Kngine: Kngine stands for Knowledge Web Search Engine intended to give significant search results by doing the study of meaning in language about the keywords and concepts, answering the questions of the users, finding the relation between the concepts and keywords and linking differeny type of data together for example images, subtitles of movies, movies, cost of the products at stores, user feedback and reviews and stories that influence. In kngine query items are isolated into one or the other web results, or picture results. They are gone before by data about the search term, known as 'Concept.' for instance, looking for the "iPhone 3GS" will be gone before by the gadget's details. Looking for a film will be gone before by data about the film, connections to trailers, surveys and statements. Looking for a city will be gone before by data about the city, neighbourhood attractions, occasions, climate and restaurants. Kngine right now contains in excess of 8,000,000 Concepts this is the place where the site's solidarity lies.

[4] Google Search Engine: Google web search is highly involved search on the planet. Google has developed a priority rank concept of 'Page Rank' and is purely implemented by the Search Engine. It separates every site by rating from 10 appearance the genuine way of behaving of the site in search engine. During page ranking many factors are taken into consideration. Among all the web browsers on the internet Google Search Engine provides more customized options and search results. It connects all the Google's product by using a single Gmail Address and gives it highest number of Extensions. Google provides image search by including special integrated tool to search GIF. It allows you to make payments online as well as notifies about the pending bills. One can can check the schedule of the flights, trains along with the time of arrival and delayed time. Package tracking is possible if the tracking number of UPS or FedEx is known. Google carry out fast parallel processing as it runs on a distributed network of thousands of low cost computers. It consist of three parts , the GOOGLEBOT which is a webcrawler that fetches and finds the web pages, then for sorting each and every word on every web page and storing the resulting index of words in a large database google has INDEXER and The query processor, which looks at your search query to the file and suggests the archives that it thinks about most significant.

[5] DuckDuckGo: DuckDuckGo is a search engine that follows the definition " The search engine that doesn't track you". The aim of DuckDuckGo is to keep the user activity private. It provides instant answers to the user query. Also the user can develop its own answers and post them just by logging the link. DuckDuckGo allows user to directly view the profile of the people on social media by putting query '@person_name' without actually leaving the search engine. DuckDuckGo gives an extraordinary element to looking the applications which will be shown as a card. Whenever you click on the card, it will give you full depiction. DuckDuckGo gives the component of Expanding and Shortening of the URLs. The main element that DuckDuckGo gives is the secret key age. For instance, if you need to make an arbitrary secret key that is not difficult to recall and difficult to break, you can look „password 16". The number can be changed, contingent upon the client prerequisite. As DuckDuckGo doesn't has paging system there is no need for the user to switch pages.

III. PROPOSED SYSTEM

- Simple User interface is provided for the user so that the complex process at backend is completely hidden from the user which makes it easy to use.
- With these UI users can enter a query and submit it.
- This query is then processed and gives away keywords.
- Through these keywords, matching results are searched into index DB.
- Based on matching keyword links of that web page is obtained.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

- In SERP (search engine result page), all these obtained pages are listed in the form of links through which users can easily navigate from one web page to another by backtracking.
- Users can easily identify the links which are most relevant to their searches because most of the links consist of keywords which are matching to the query.
- Even a search engine helps users to give the best result at the top and this best result is optimized in such a way that a link should contain complete keywords at the top.

IV. METHODOLOGY

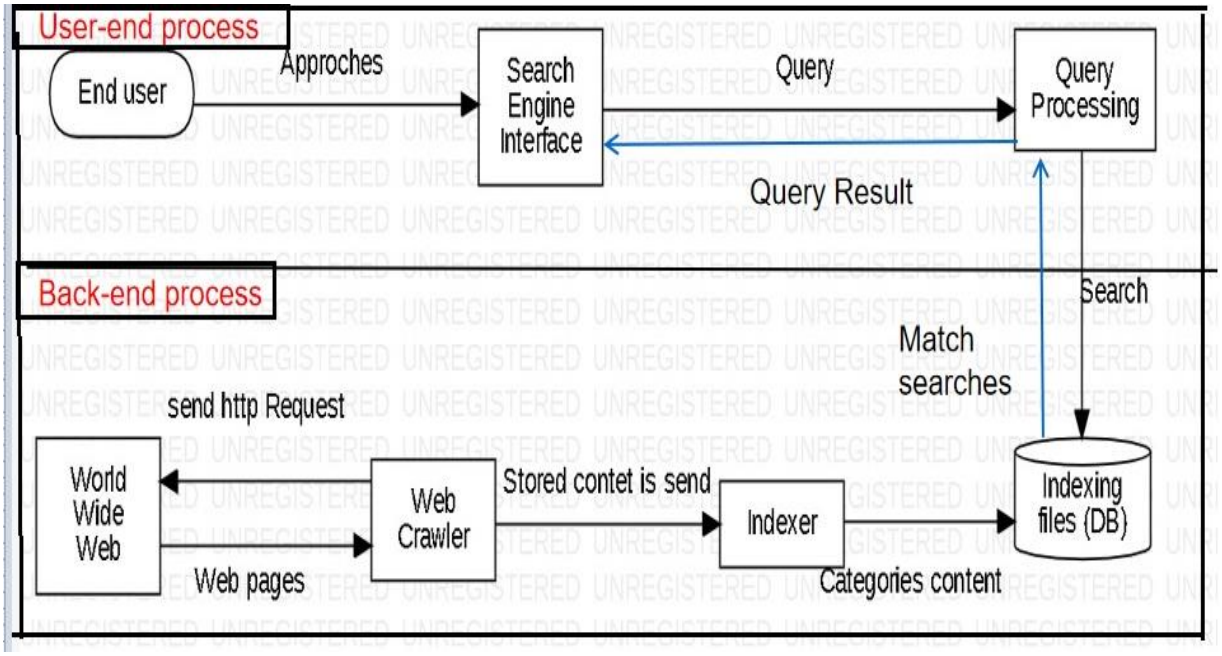

Fig 1: Architecture Diagram for search Engine

As you can see in the architecture, how the system components will interact with one another from backend to end-user. So, in backend web crawler is design which is a internet bot/spider or software that browse the world wide web(www) in systematic way by visiting web pages and parse for further indexing process. While, visiting website web crawler send https request to www through link web crawler has gathered and in return gets web page to crawl their content and download it and send to indexer while crawling extract links are kept with crawler for further crawling then indexer categories content in huge database and when user fires query is goes through query processing or engine to match the content of indexer based on keywords to return information in the form of results.

There are 3 main process to carried out behind search engine simultaneously:

1. Crawling: -Crawling is the process of discovering web pages through URLs. Search Engine sends out a team of robots (known as crawlers or spiders) to find Web pages with URL links available on pages.

- The first step is finding out what pages exist on the world wide web.
- A web crawler starts with a list of URLs to visit, called seeds. It will visit the web pages and copy or download it for processing.
- Web pages are discovered by following links from a known (already visited) page to a new page.
- Search engines must constantly search for new pages and add them to its list of known pages. This will help to update web content daily.
- Information obtained by crawler is stored in the repository of web pages.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

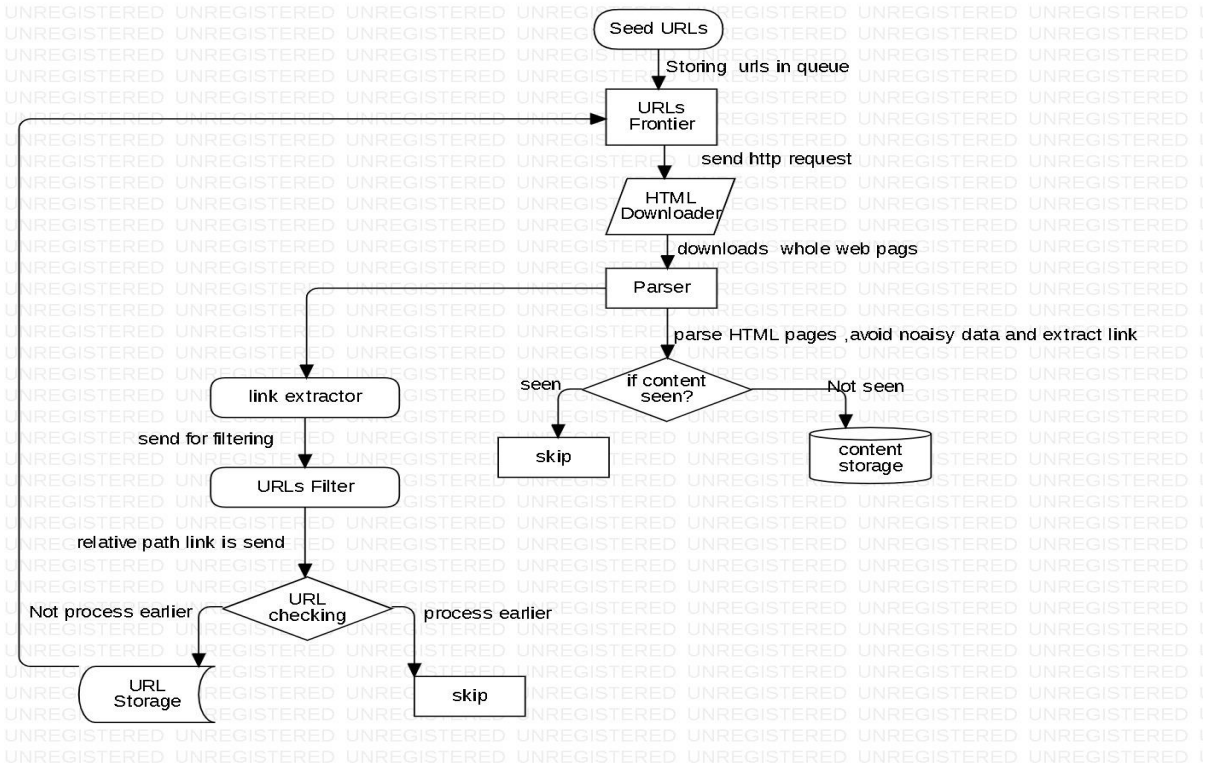

Fig 2: working of web crawler with parser

Initially Seed URL is served to crawler one thing to keep in mind while choosing seed URL is it should contain lot many links. Then send https request and parser is responsible for checking whether contents is already parsed or not if parse then skip that contents move to next and the another task as parser is to extract links it also takes care that no same links are parsed. Links which is not process earlier is then store in URL storage and these is send to URL frontier for further crawling.

Parsing: -Parsing allows transforming data in a way that can be understood by a specific software. This process can be associated with indexing but here we are using it with crawling. After downloading full content of website parsing comes into picture. It will look for text and non-text type data and categories it in two parts:

1) parsing takes the visible text portion from downloaded content and send further for indexing

2) It pulls out the links from downloaded content for future crawling. This process will work recursively till the entire web is not crawled. Crawling and parsing both this process works smoothly and gives result in efficient way

3) Indexing: -After a page is discovered and parse, search engines try to understand what the page is about. This process is called indexing. It will look through the visible text portion pulled out from parsing for further analyzing and understanding the content. After learning the content (i.e., what the page is about) the meaningful information is extracted out it is sorted, organized, and stored in a huge database. Once a page is in the index, it's good enough to serve up to searches.

Features of search engine

- Ranking System: -Through ranking we can achieve the nearest correct result for user query. Firstly, we rank the result based on the word individuality is at the top means the word which contain space around it is at top rather than substring of the word. But as this method as positive as well as negative effect on other ranking result as we said it should contain space around it so it will give pages at top which as space before word rather than actual keyword which is at start. So, second ranking factor is that when a keyword is found at the start of the title then it is ranks high. Main domain web pages should be at the top and the pages which refers this domain should be after it, in order wise manner has in list of links. It is mostly dependent on index.
- Autosuggestion: -We have also implemented auto suggest function in our search engine but improves user

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

experience. These automatically suggests or predicts the next letter word which user is trying to type.

- Highlights: - This feature is used to highlights the user query with color in the search result. So, that user will be able to gaze it.

V. RESULTS

Simple User Interface with search engine logo and search bar to search anything

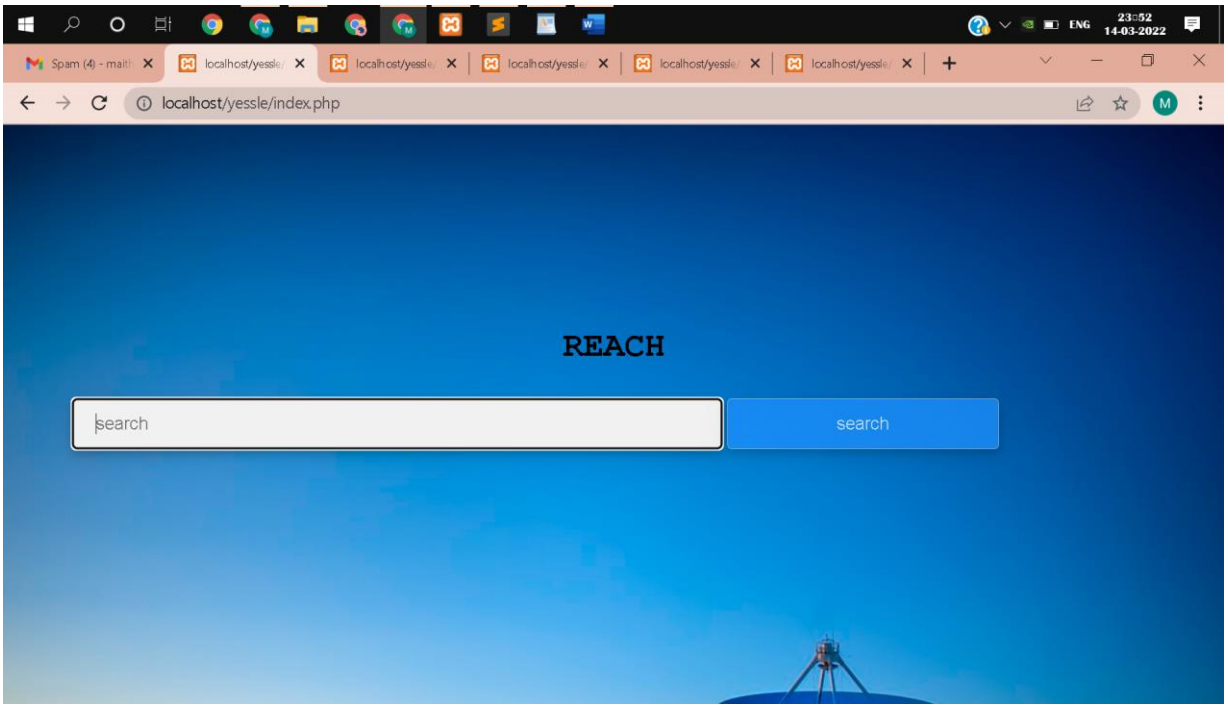

Fig 3: Home Page

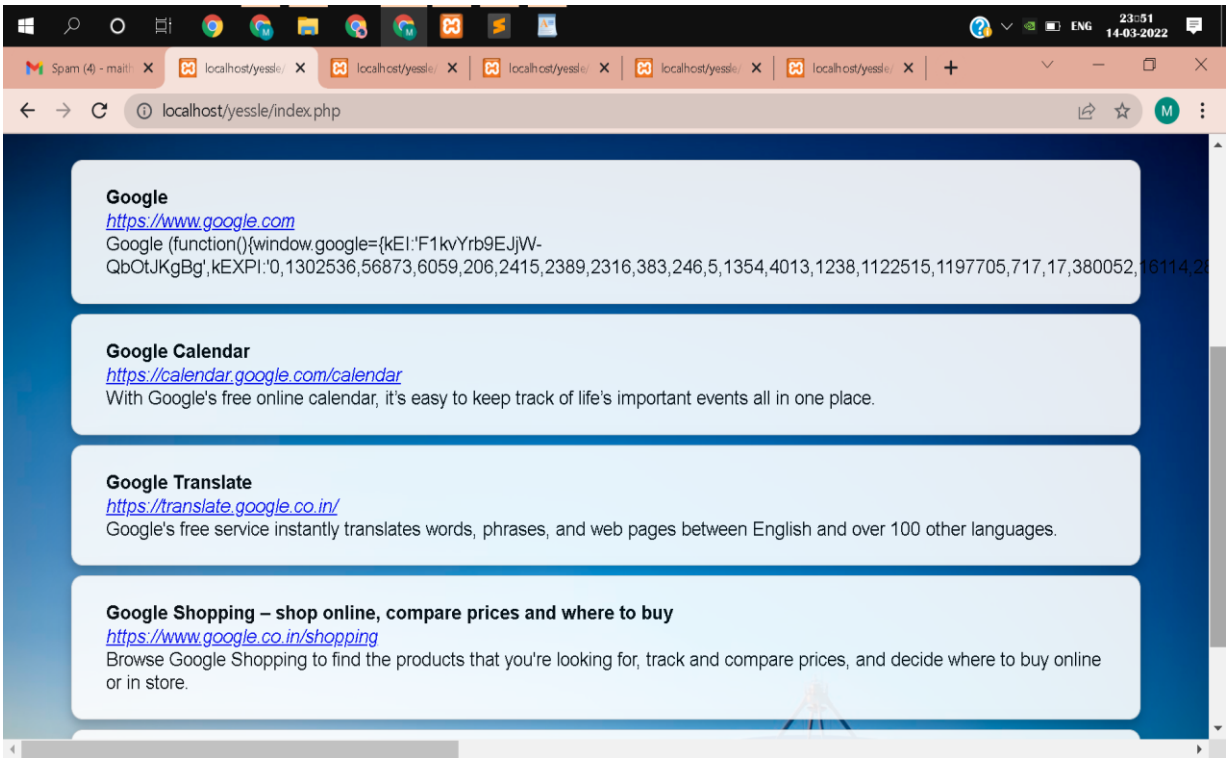

Fig 4: SERP

Query related information is fetched in the form of links to that website in which information is present on search result page.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

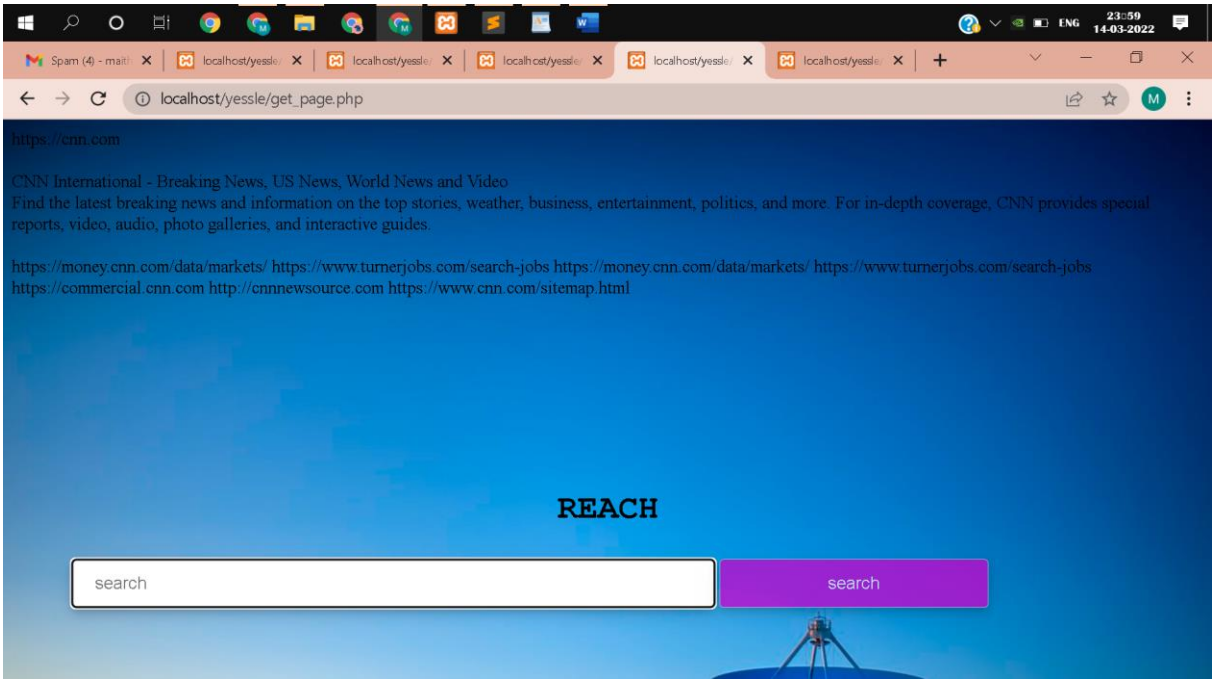

Fig 5: Index Engine

When web address is given to index engine is send http or https request to server demanding for webpage. Then it extracts html content from title tag, links and description and store it in database. Which is further user by query engine to retrieve information according to user query.

VI. CONCLUSION

This paper describes the brief overview of our search engine working and different approach to find result according to user query. It is observed that searching the internet today is a challenging because of lot more information is available and uses the same keyword but in different way which was quit tough to give relevant result to user which they were looking for, so half of query which is complex go unanswered but most of it is estimated. But ranking factor actually improve the probability of finding relevant results and further it can be improve by introducing more ranking factor .

ACKNOWLEDGMENT

I would like to convey my special thanks of gratitude to project guide professor and project coordinator of Information Technology Mrs. Smita Pai who gave us the good opportunity to do this amazing project on the topic search engine which also helped us in doing a lot of research and providing us with all the resources that was required for this project.

VII. REFERENCES

[1] Wang, H.; Tan, C. C.; Li, Q. (2008). [IEEE IEEE INFOCOM 2008 - IEEE Conference on Computer Communications - Phoenix, AZ, USA (2008.4.13-2008.4.18)] IEEE INFOCOM 2008 - The 27th Conference on Computer Communications - Snoogle: A Search Engine for the Physical World., (), 1382–1390. doi:10.1109/infocom.2008.196

[2] Staab, Steffen; Svátek, Vojtěch (2006). [Lecture Notes in Computer Science] Managing Knowledge in a World of Networks Volume 4248 || SemSearch: A Search Engine for the Semantic Web, 10.1007/11891451(Chapter 22), 238–245. doi:10.1007/11891451_22

[3] Ostermaier, Benedikt; Romer, Kay; Mattern ,Friedemann ;Fahrmair, Michael; Kellerer, Wolfgang (2010). [IEEE 2010 Internet of Things (IOT) - Tokyo, Japan (2010.11.29-2010.12.1)] 2010 Internet of Things (IOT) - A real-time search engine for the Web of Things, (), 1–8. doi:10.1109/IOT.2010.5678450

[4] Semantic Web Search Engines: A Comparative Survey January 2019 International Journal of Scientific Research in Computer Science Engineering and Information Technology DOI:10.32628/CSEIT195115

[5] https://www.researchgate.net/publication/221035602_Design_and_Implementation_of_a_Geographic_Search_Engine

[6] Romer, K., Ostermaier, B., Mattern, F., Fahrmair, M., & Kellerer, W. (2010). Real-Time Search for Real-World Entities: A Survey. Proceedings of the IEEE, 98(11), 1887–1902. doi:10.1109/jproc.2010.2062470

[7] K.-K. Yap, V. Srinivasan, and M. Motani, BMAX: Human-centric search of the physical world,[in Proc. 3rd Conf. Embedded Netw. Sensor Syst., 2005, pp. 166–179.

[8] T. Yan, D. Ganesan, and R. Manmatha, BDistributed image search in camera sensor networks,[in Proc. 6th Conf. Embedded Netw. Sensor Syst., 2008, pp. 155–168.

[9] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti, BCellular census: Explorations in urban data collection,[IEEE Pervasive Comput., vol. 6, no. 3, pp. 30–38, Jul./Sep. 2007.

[10] Microformats. [Online]. Available: http:// microformats.org

[11] https://www.javatpoint.com

[12] https://www.w3schools.com

[13] https://www.tutorialspoint.com

[14] "DuckDuckGo search engine", Internet:http://www.duckduckgo.com, [July, 2016]

[15] "Features of DuckDuckGo", Internet:http://www.duckduckhack.com/, [July, 2016]

[16] "Features of Google search Engine", Internet: http://thenextweb.com, [July, 2015]

[17] "Architecture of DuckDuckGo", Internet: http://insightsdelight.wordpress.com, [August, 2016]

[18] "Architecture of Google Search Engine", Internet: http://googleguide.com, [August,2016]

[19] M. Cui and S. Hu, "Search Engine Optimization Research for Website Promotion," 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, Nanjing, Jiangsu, 2011, pp. 100-103.

[20] "DuckDuckGo vs. Google Search Engine",Internet:http://www.maketecheasier.com, [August, 2016].